



**LAHDEN AMMATTIKORKEAKOULU**  
*Lahti University of Applied Sciences*

# OMSTRUKTURERING AV ORTNAMNSREGISTRET

En förstudie

LAHTI YRKESHÖGSKOLA  
Teknik  
Informationsteknik  
Programvaruteknik  
Examensarbete  
Våren 2014  
Jan Paakkanen

Lahti yrkeshögskola

Utbildningsprogrammet för informationsteknik

PAAKKANEN, JAN:

OMSTRUKTURERING AV  
ORTNAMNSREGISTRET  
En förstudie

Examensarbete för programvaruteknik, sidantal 44

Våren 2014

SAMMANDRAG

---

Detta examensarbete berör ISOF (Institutet för språk och folkminnen) i Uppsala och dess Ortnamnsregister med olika databaser. Man planerar att omstrukturera databaser hos ISOF. Därför skulle det vara bra att veta vad användare av Ortnamnsregistret har för förslag på förändringar och vad man överhuvudtaget behöver förändra i databassystemmiljön. Man borde skapa sig en realistisk koppling mellan behov och möjligheter.

I början skildras Ortnamnsregistrets IT-miljö och ortnamnforskningen, hur och i vilken mån de påverkar varandra. Två stora forskningsområden är etymologi och befolkningshistoria.

Man har då gått genom den tekniska sidan av databassystemet i Ortnamnsregistret. I centrum står MySQL och PHP.

Sedan ser man olika möjligheter att förändra IT-miljön. Man borde uppdatera en del av systemet, för man använder många åldrande mjukvaruversioner. Några förslag är svåra att genomföra, de är för dyra, har kanske bara begränsad nytta osv. men i en modifierad form behöver man bara belysa några orter noggrannare eller generera det som behövs på ett annat sätt än man gjort förr och man kan gå vidare. Det finns också mindre förslag som går ihop med större förändringar om de förverkligas, förändring av fonter till exempel.

I slutet presenteras en åtgärdsplan. Planen är mer konservativ än innovativ.

Nyckelord: databas, ortnamn, PHP, MySQL

Lahti University of Applied Sciences  
Degree Programme in Information Technology

PAAKKANEN, JAN:

Restructuring a place name registry  
Preliminary study

Bachelor's Thesis in Software Engineering, 44 pages

Spring 2014

ABSTRACT

---

This thesis deals with Ortnamnsregistret, the place name registry of ISOF, Sweden's language and folklore institute, in Uppsala, Sweden. There is a plan to restructure the databases of the registry. In connection to that, this would be a proper time to find out what kind of changes the users might need or want and what in general needs to be changed. There should be a realistic connection between the needs and the possibilities. This thesis is a preliminary study, to help the restructuring process.

First, there is an outline of how the Ortnamnsregistret operates and the IT environment connected with it, focusing on how the IT environment is used for research purposes. The two key research areas are etymology and population history.

Then the technical side of the database system of Ortnamnsregistret is described. The focus is on MySQL and PHP.

Then different areas in the IT environment that need changing and the possibilities to do that are considered. An update is needed since some of the software is quite old. Some suggestions are difficult to implement; they are too expensive, are of little use, and so forth. Nevertheless, it would be feasible to update those places that are considered more important. Instead of using a difficult process to feed data to a database and random search for information, the process could be simplified by using XML unit files and lists. There are minor changes, such as having new fonts, which are done in connection to more important changes, if those are put into effect.

At the end, a concrete plan of action is offered. The plan is fairly conservative, rather than innovative.

Key words: database, place names, PHP, MySQL

## INNEHÅLL

1	INLEDNING	1
1.1	Bakgrund	1
1.2	Forskningsproblem	2
2	ORTNAMNSREGISTRET	3
2.1	Ortnamnsregistret, ortnamnsdatabaser och IT	3
2.2	Ortnamnsregistret och ortnamnsforskning	4
2.3	Andra användare	7
3	NUVARANDE DATABASSYSTEM	8
3.1	Databasstruktur	8
3.1.1	Andra databaser	11
3.2	Mjukvara	14
3.2.1	Uppdateringspolitik	16
4	TANKAR OM FÖRÄNDRINGAR OCH FÖRSLAG	17
4.1	Databasförändringar	17
4.1.1	Beläggsår	18
4.1.2	Etymologisk forskning	19
4.2	Hårdvara	21
4.2.1	Uppdateringspolitik	21
4.3	Mjukvara	21
4.3.1	Inmatningsprogram	21
4.3.2	Sökprogram	24
4.3.3	MySQL	25
4.3.4	PHP	27
4.3.5	Testfall	28
4.4	Andra förändringar	29
4.4.1	Fonter	29
4.4.2	Teckenkodning	32
4.4.3	Att korrigera fel	34
4.4.4	Loggande, när och var någonting har gjorts	34
5	ÅTGÄRDSPLAN	36
5.1	Databasförändringar	36
5.1.1	Att generera listor av ortnamn på olika skäl	36

5.2	Mjukvaru- och uppdateringspolitik	37
5.3	Inmatningsprogram	37
5.4	Sökprogram	38
6	SAMMANFATTNING	39
	BIBLIOGRAFI	41

## 1 INLEDNING

ISOF (Institutet för språk och folkminnen) är ett institut som ansvarar för officiella språkliga tjänster och forskning anknuten till dessa. Inom ISOF arbetar ett tiotal forskare som forskar i ortnamn, personnamn och övriga namn, svenska dialekter, folkminnen, de svenska nationella minoritetsspråken i Sverige osv. Språkrådet med sin rådgivande verksamhet i svenska och de nationella minoritetsspråken ingår också i ISOF. Detta arbete kommer att rikta sig mot ISOF i Uppsala, och i synnerhet mot Ortnamnsregistret där. Ortnamnssamlingar i Ortnamnsregistret är största i världen. (Institutet för språk och folkminnen 2014.)

Detta arbete börjar med att ta reda på vilka behov Ortnamnsregistrets och de andra ortnamnsdatabasernas användare, först och främst ISOFS forskare, har och om IT-avdelningen kan hjälpa till, d.v.s. vilka behov är mer eller mindre möjliga att fylla. Man måste givetvis begränsa arbetet, till exempel berör man egentligen inte nätverksfrågor. Som sagt kommer man huvudsakligen att begränsa sig till ISOFS Ortnamnsregister i Uppsala (och till andra ortnamnsdatabaser hos ISOF), även om något kan ha en vidare användning. Detta arbetes tyngdpunkt ligger på ortnamnsdatabaser och sökfunktioner.

### 1.1 Bakgrund

Under hösten 2014 har man planerat att börja en omstrukturering av ISOFS olika databaser. ISOF har databaser i formaten FileMaker, MySQL och XML (och några andra mindre databaser). Man skulle vilja ha en enhetlig hantering av de här databaserna och naturligtvis göra de förändringar som behövs och önskas.

Det är önskvärt att förändringar planeras på förhand och att planeringen är gjord med genomtänkta, lagom stora och innehållsmässigt passande bitar. I detta sammanhang i detta arbete är Ortnamnsregistrets MySQL-databas över Sverige i fokus, men andra saker kommer att dyka upp, till exempel omständigheten att man har databaser i andra format hos ISOF såsom XML-databaser.

## 1.2 Forskningsproblem

Det är ofta svårt att få experter inom olika discipliner att förstå varandra. Det är också en orsak till att vanliga konsultrapporter inte alltid är det som behövs.

Därför är det kanske bättre att ha någon som har erfarenhet gällande de olika disciplinerna för att försöka förmedla mellan disciplinerna. Här kommer personlig historia och erfarenhet fram. Och det är också en orsak att detta arbete nu har kommit till stånd.

Det centrala forskningsproblemet här är hur man ska utveckla ortnamnsdatabasen inom den närmaste framtiden. Men med ortnamnsdatabasen kommer mera saker fram som man bör tänka på. ISOF erbjuder sina forskare en arbetsmiljö med rum, dator osv. Det är ganska lätt att begränsa miljön som är relevant här till mjukvaran och hårdvaran. Men inom mjukvaran och hårdvaran kommer vissa problem fram. Om några browserversioner har problem med sökningen i ortnamnsdatabasen så är det ett IT-problem. Det är också inte lätt att korrigera problem som kommer från att man har en åldrande servermiljö, ifall det är en utomstående firma som sköter denna servermiljö. Ett lite bredare perspektiv är viktigt också här, för något utanför ortnamnsdatabasen och sökfunktionerna kan verkligen hindra användbarheten av Ortnamnsregistret.

Kapitel två handlar om Ortnamnsregistret. I kapitel tre skildras det nuvarande databassystemet. Kapitel fyra är analysen. Kapitel fem innehåller åtgärdsplanen.

## 2 ORTNAMNSREGISTRET

### 2.1 Ortnamnsregistret, ortnamnsdatabaser och IT

Ortnamnsregistret består av en MySQL-databas, med drygt 3,7 miljoner inskannade och registrerade sedeskort, och av program som gör något med eller för databaser. Databaser är samlingar av data med någon struktur därinne. Idén bakom olika databaser är också olik: strukturerade och delade enheter, tabeller med kolumner i relationella databaser och mer eller mindre ett datalager, ”en enda tabell”, med struktur i XML-databaser. Av erfarenhet kan man säga att relationer i IT-sammanhang inte är speciellt viktiga gällande ortnamnen, därför är en flat struktur och en datalagerlösning d.v.s. XML-format lockande.

Databasen i Ortnamnsregistret hanteras först, men inte kanske främst, med MySQL. Man har i princip bättre möjligheter att söka saker i en databas med MySQL i kommandoraden än med mera raffinerade lösningar, till exempel kan man söka på nytt med resultatet från den första sökningen. Dock kommer användare vanligen inte att söka saker på detta sätt, för

- datan i sig inte är så komplicerad eller i en (komplicerad) form som tillåter komplicerade sökningar, och mera komplicerade aspekter av sökningar går att göras för hand
- vanliga användare kommer inte att lära sig SQL-syntax
- det finns få orsaker att söka komplicerade saker (det är svårt att tänka att forskare drar större nytta av komplicerade SQL-sökningar för till exempel komplicerade SQL-sökningar använder relationer mellan tabeller, och om det finns få tabeller försvinner nyttan).

XML-formatet erbjuder lite annorlunda erfarenheter (kanske lite mer användarvänliga för lekmannen), men i princip samma saker som sagts om MySQL gäller här också.



## 2.2 Ortnamnsregistret och ortnamnsforskning

Ortnamnsforskningen är vanligen datacentrerad nuförtiden skaffar man datan först och främst genom databaser men vid behov även från informanter. För att uppnå några mål med databaserna, måste man a) ha en uppfattning vad för data som finns, b) till vad man använder datan och c) under vilka villkor man kan uppnå sina mål.

Ur IT-avdelningens synvinkel kan man tänka sig att många forskares behov kanske inte är så intressanta och att de någon gång nästan är på snudden till banala, till exempel sökningar som man redan kan göra med sökprogrammet. Icke desto mindre är det centrala med databaser sökmöjligheterna. Ett datalager är väl värdelöst utan användning. Man kan alltid göra specialiserade sökfunktioner men den här vägen är oändlig. Om man vet hur man söker i databaser med SQL, så behöver man egentligen inte specialiserade sökfunktioner. Ändå kan man inte kräva att forskare använder tid till att lära sig SQL, tyvärr. Det finns dock förslag som är svåra att implementera, och då måste man fråga sig om de även är värda att implementera.

För lekmannen kan det kännas att ortnamnsdatabaser inte ofta är så viktiga. Det som har sagts kan också tolkas så (Ainiala & Pitkänen 2002, 87):

*EERO KIVINIEMI discussed the settlement historical document value of place names in principle in his article "Nimistö Suomen esihistorian tutkimuksen aineistona" (Names as research material of Finnish prehistory; 1980). He evaluated the possibilities of existence and identification in the present Finnish stock of the old loan names from different languages — Baltic, Proto-Germanic, Proto-Scandinavian and Lappish. KIVINIEMI presented many causes of uncertainty, related to settlement historical conclusions drawn on the basis of old place names. For example, the contents and the factual background of the names often remain ambiguous, or they may not be accessible at all, which is why drawing direct conclusions on the basis of distribution and typological criteria is often misleading. KIVINIEMI emphasized that it is only possible to draw reliable settlement historical conclusions on the basis of extensive data from basic research into place names. He deals with issues of the same type in his article "Namenforschung und Siedlungsgeschichte in Finnland" (1981).*

Men även om ortnamnsforskningen kräver något mer än ortnamnsdatabaser, och ofta är det så, kan inte de där andra källorna vara så omfattande och exakta som ortnamnsdatabaser som går rakt på sak. Man kan nog ha en kvasi-exakt bild om någon sak i ortnamnsforskningen utan ortnamnsdatabaser, men andra källor är inte först och främst ortnamnskällor, och forskning utan ett omfattande basmaterial riskerar att stanna som kuriosa. Egentligen är det utan tvivel ett plus om man har ortnamnsdatabaser och andra källor (skattearkivet osv.), men även om ortnamnsdatabaser inte erbjuder länkar till mycket annat än ortnamnsmaterial, så kan andra källor vanligen erbjuda godtyckligt eller inte relevant material. Alltså är ortnamnsdatabaser väsentliga i den systematiska ortnamnsforskningen, allt annat är ett plus.

Det kanske hjälper att ta med ett exempel för att försöka förklara varför det är så. Vikstrand har ett väldigt illustrativt exempel med ortnamnet Vedyxa (Vikstrand 2008). I figur 1 kan man se hur ortnamnet tolkades annorlunda av olika nedtecknare samtidigt som det förändrades under historiens gång (Vikstrand 2008, 466).

j vidhusum 1358  
 j widhusum 1397  
 i Widhusom 1403  
 i Vidhusum 1415  
 j widhwsa 1425  
 j vidhusom 1426  
 vidhussom 1428  
 j widhusom 1430  
 j vidhøssa 1440  
 i vidissiø 1456  
 i widwsä 1492  
 Vedyse, Vedesse 1540  
 Weddessö, Vedzö 1542  
 i vedyssä 1554  
 Widessö, Wedessa 1555  
 Vedhyxe 1561  
 Wedyxa 1581  
 i Wedyxa 1591

Figur 1. Förändring av ortnamnet *Vedyxa* (Vikstrand 2008, 466)

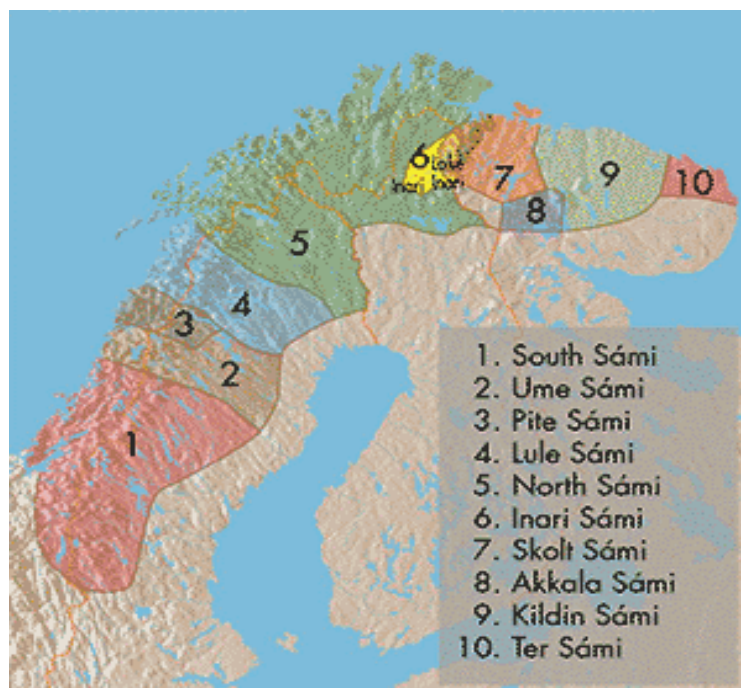
I Vikstrands artikel är Vedyxa ett exempel på att folketymologi genom att göra ortnamn förståeliga skapar ordning i kaoset som ljudförändringar tillbringar. Det

är knappast så att forskare går igenom Ortnamnsregistret och hittar ortnamn som leder genast till artiklar som Vikstrands, men utan sådana möjligheter är det omöjligt att även börja forska i ortnamnet. Det finns nog arkiv, men att hitta något där är slumpmässigt. Det är kanske svårt att komma på det man ska söka, men för att eventuellt hitta något, har man bättre möjligheter med ortnamnsdatabaser för de fokuserar på frågan. Vikstrand säger i artikeln att (Vikstrand 2008, 466):

*Ved- kan visserligen förekomma som förled, men -yxa är definitivt ingen etablerad ortnamnsefterled. Man kan därför inte påstå att ombildningen gjort namnet mer »ortnamnslikt».*

Man kan nämligen göra ortnamnslistor, och även baklängesortnamnslistor samt listor på en godtycklig grund. Med strukturella elektroniska databaser är ett sådant arbete mycket lättare och överhuvudtaget mera omfattande än någonsin förr. Det är inte möjligt att ge en formell definition av ”ortnamnslik”, men som den första approximationen kan man ta ut transparenta ortnamn. Det som återstår kunde vara en källa eller ett uppsving för forskningen. Det finns ändå mycket att göra innan man skriver artiklar och böcker, men man ser dock att med hjälp av det som IT har att erbjuda, är möjligheterna att skriva de där artiklarna och böckerna överhuvudtaget mycket mer sannolika.

ISOFs forskare forskar också i befolkningshistoria. Det finns någonting mer att säga om saken än vad är allmänt trott. Låt oss ta ett exempel på samernas boställen. Den allmänna bilden är med små variationer nästan alltid den som kan ses i figur 2 (Morottaja 2006) som visar utbredningsområdena för de samiska språken kopplade till de samiska sommar-, vår-, höst- och vinterbetesområdena ett vanligt år, reservbetesområdena är inte utmärkta. Men man vet att detta inte är hela sanningen. Man känner till exempel till sockenlapparna och var de levde (Svanberg 1986, 103). Sockenlapparna hade slutat att vara en levande etnicitet på 1800-talet (Svanberg 1986, 97) före systematiska ortnamnssamlingar (Namnarkivet i Uppsala 2010); man vet att spåren av minoriteter överhuvudtaget är få inom majoritetssamhället, och att samla in saker direkt från ett folk självt är naturligtvis säkrare än att använda ett annat folk som källa. Icke desto mindre om man har omfattande ortnamnssamlingar kommer spåren av andra folk sannolikt att dyka upp i materialet.



Figur 2. En ganska traditionell samisk språkkarta (Morottaja 2006)

### 2.3 Andra användare

Förutom forskarna kunde administrationen möjligen ha några behov. Någon sorts övervakning är typiskt för administrationens arbete. Forskningen är typiskt svår att övervaka, vilka är relevanta kriterier för övervakandet. Troligen har administration dock inte så stora behov.

Vanligt folk som använder Ortnamnsregistrets databas erbjuds i princip samma sökfunktioner som forskarna. Att granska i fulla vad folk i allmänhet vill göra med Ortnamnsregistret går utanför detta arbetes ramar, men det finns inga stora krav så långt ISO vet. Utan tvivel finns det amatörer som är väldigt intresserade av ortnamn, några kanske vet hur man söker ordentligt med SQL osv. Troligen kommer man i framtiden att ge lekmanen mera möjligheter att använda ortnamnsdatabaser eftersom trenden inom språkbranschen är mot mera öppenhet, till exempel Svenska Akademiens ordlista är för alla åtkomlig (Svenska Akademien 2006).

### 3 NUVARANDE DATABASSYSTEM

Här beskrivs det centrala databassystemet inom Ortnamnsregistret. En firma sköter ISOFs server-hårdvara och mjukvaran i servern är lite gammal och med tiden blir det svårare och svårare att uppdatera den. Den mjukvara som man använder för att utnyttja databasen beskrivs här. Systemet kommer att beskrivas närmare och bredare och genom att göra det så berättar det för oss något speciellt om Ortnamnsregistret och dess upprätthållande.

#### 3.1 Databasstruktur

Databasen följer den administrativa indelningen som existerade i Sverige 1951. Databasen *ortnamndb* innehåller tabeller *ortnareg* och *sockensida* (Skånsk ortnamnsdatabas är lite annorlunda) vars CREATE CLAUSE är följande (figur 3 och figur 4). Innehållet av tabellernas kolumner skildras nedan: många tabeller skildrar den administrativa indelningen, ett par skildrar ortnamntypen, några är byråkratiska och hjälper med databasstruktureringen, några är tomma eller nästan tomma.

```

1 CREATE TABLE IF NOT EXISTS `ortnareg` (
2   `alan` char(2) NOT NULL,
3   `akod` char(7) NOT NULL,
4   `aurv` char(1) NOT NULL,
5   `akat` char(2) NOT NULL,
6   `adok` char(8) NOT NULL,
7   `alfk` varchar(19) NOT NULL,
8   `aort` varchar(50) NOT NULL,
9   `asok` varchar(50) NOT NULL,
10  `arub` varchar(50) NOT NULL,
11  `alok` varchar(50) NOT NULL,
12  `ahrd` varchar(50) NOT NULL,
13  `alkp` varchar(14) NOT NULL,
14  `atol` char(1) NOT NULL,
15  `abea` varchar(10) NOT NULL,
16  `abeb` varchar(512) NOT NULL,
17  `atxt` varchar(2000) NOT NULL,
18  `apub` date NOT NULL default '0000-00-00',
19  `anum` char(2) NOT NULL,
20  `arad` decimal(30,25) NOT NULL,
21  PRIMARY KEY (`alan`,`akod`,`aurv`,`akat`,`adok`),
22  KEY `asok_Index` (`asok`),
23  KEY `aort_Index` (`aort`,`ahrd`),
24  KEY `aort2_Index` (`aort`,`ahrd`,`alfk`)
25 ) ENGINE=MyISAM DEFAULT CHARSET=latin1;
```

Figur 3. CREATE CLAUSE för tabellen *ortnareg*

Tabellens innehåll kan beskrivas på följande sätt:

- alan, länet
- akod, socknens kod
- aurv, om bilden är en FramSida eller BakSida
- akat, bebyggelse eller naturnamn
- adok, bildens filnamn
- alfk, länets namn
- aort, ortnamn på socken/stad/köping/municipalsamhälle
- asok, socknens ortnamn
- arub, rubrik d.v.s. det egentliga ortnamnet
- alok, lokaltypen
- ahrd, härads/skepplags/bergslags/tingslagsnamn
- alkp, landskapsnamn
- atol, om tolkning av rubriken finns
- abea, första beläggsåret
- abeb, eventuellt flera beläggsår
- atxt, ytterligare information och kommentarer
- apub, datum då posten eller bilden publicerats på webben
- anum, nummerordning för länets alfk
- arad, radnummer på socken-HTML-sidan

```

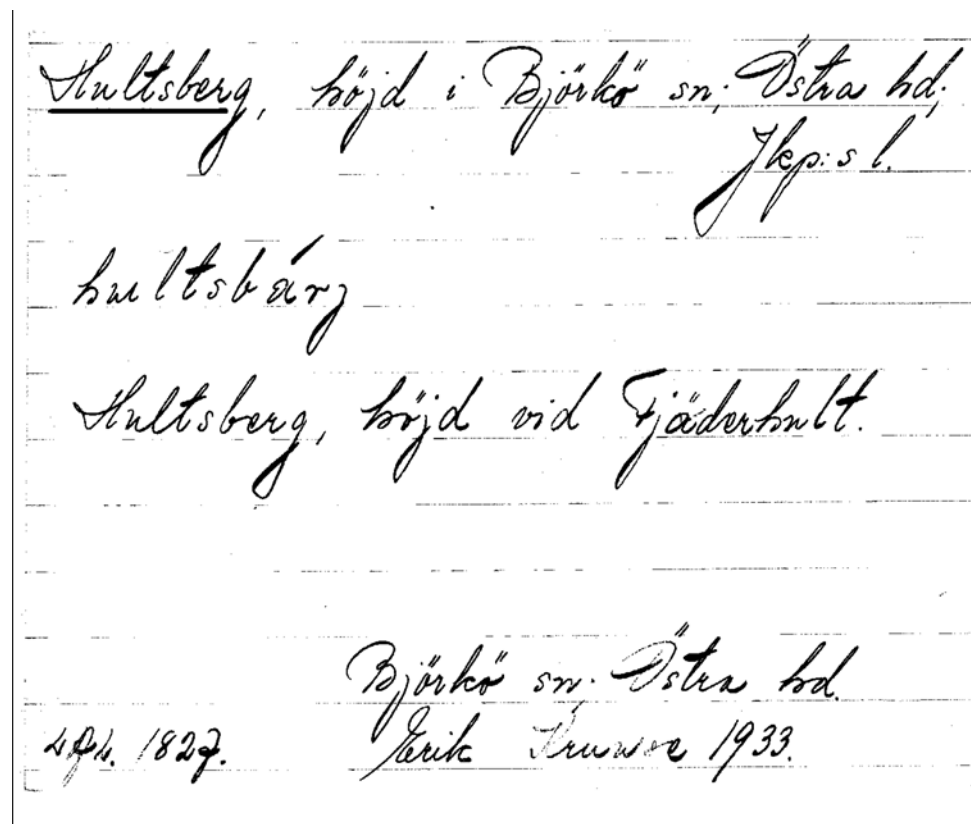
1 CREATE TABLE IF NOT EXISTS `sockensida` (
2   `alan` char(2) NOT NULL,
3   `akod` char(7) NOT NULL,
4   `aort` varchar(50) NOT NULL,
5   `agen` varchar(54) NOT NULL,
6   `alkp` varchar(14) NOT NULL,
7   `alfk` varchar(19) NOT NULL,
8   `ahrd` varchar(50) NOT NULL,
9   `ahad` varchar(70) NOT NULL,
10  `arad` varchar(5) NOT NULL,
11  PRIMARY KEY (`alan`,`akod`),
12  KEY `iasok` (`alan`),
13  KEY `ialok` (`akod`)
14 ) ENGINE=MyISAM DEFAULT CHARSET=latin1;
```

Figur 4. CREATE CLAUSE för tabellen *sockensida*

Tabellens innehåll kan beskrivas på följande sätt:

- alan, länet
- akod, socknens kod
- aort, ortnamn på socken/stad/köping/municipalsamhälle
- agen, en tolkad version av aort
- alkp, landskapsnamn
- alfk, länets namn
- ahrd, härads/skeppslags/bergslags/tingslagsnamn
- ahad, nästan alltid precis samma som ahrd, 7 undantag
- arad, radnummer på socken-HTML-sidan

Man har strukturerad information om ortnamnen i databasen så långt det går. Men för att kunna ha mer, har man också skannat in de ursprungliga anteckningarna, se ett exempel i figur 5. Informationen om var ortnamn ligger är övertagen till Ortnamnsregistret, fast allt annat visar man inte strukturellt i Ortnamnsregistret.



Figur 5. Ett exempel på ortnamnsanteckningarna, Hultsberg

### 3.1.1 Andra databaser

ISOF har andra databaser vars struktur är annorlunda. Ofta behandlar de saker utanför ortnamnsanteckningar. Genom erfarenhet har antalet kolumner i databaserna minskat, och skepticism emot nya kolumner ökat (IT-avdelningen och forskarna säger så).

FileMaker-ortnamnsdatabaserna hos ISOF har många tabeller och komplicerade relationer. Vad man matar in kan vara relativt okomplicerat (figur 6), men alla tabeller och relationer är ganska omfattande (tabell 1).

Som ett exempel på en omfattande tabell kan vi ta Pers som har information över insamlarna. Tabellens 22 fält är PersId, Namn, Född, Osäker, Titel, Kön, Personalia, Antal\_acc, PostSkapad, PostÄndrad, !AnvÄndradAv, !AnvSkapadAv, Namn\_, Dublett, !Org, Namn\_Allt, !Acc, !Roll, PersIdDubblettFörsta, Namn\_Roll, VisaAntalAcc och Web\_Namn\_Allt\_Web.

XML-databaser är inte i (officiellt) bruk ännu.

**EXCERPERINGSMALL**

Uppslagsform

Lokatyp   
(avsluta med , + tomslag)

Socken   
(avsluta med , + tomslag)

Härads etc   
(avsluta med , + tomslag)

Län   
  
Beräkningsfält

Scriptfält, ange 07 (används endast när Beräkningsfältet inte är till fyllt).

Land/landskap

Etym/uttalsf/källa

Ny post

Skriv ut sedes m

Skriv ut sedes

Visa sedes

Visa lista

Visa sedes med s

Avsluta

Anteckningar

Befrämning/Änge B

Excerptist   
Sätt din signatur som förva

Ex. dat  Postar

Figur 6. Att mata in data i en FileMaker-databas



Tabell 1. FileMaker-ortnamnsdatabastabellerna

Tabellnamn	Statistik	Förekomster i relationsdiagram
Global	30 fält definierade, 1 post	<a href="#">Global Org</a> , <a href="#">Global Realkatalog</a>
Acc	103 fält definierade, 246242 poster	<a href="#">Acc Realkatalog</a> , <a href="#">AccNrTx</a> , <a href="#">Acc</a> , <a href="#">Acc Acc Grupp</a> , <a href="#">Acc Acc GruppSkapa</a> , <a href="#">Acc 2</a> , <a href="#">Acc Acc GruppVisa</a>
Form	68 fält definierade, 260512 poster	<a href="#">Form Acc Alla</a> , <a href="#">Form Acc Tidning</a> , <a href="#">Form Acc Skriftilig</a> , <a href="#">Form Acc Fil</a> , <a href="#">Form Acc Inspelning</a> , <a href="#">Form Acc Foto</a> , <a href="#">Form Acc Mikrofilm</a>
AccOrt	28 fält definierade, 254782 poster	<a href="#">AccOrt</a>
Ort	23 fält definierade, 11615 poster	<a href="#">Ort</a> , <a href="#">Ort Global LandskapVI</a> , <a href="#">Ort Ort Dubblett</a> , <a href="#">Ort Ort ByVI</a> , <a href="#">Ort Global ByVI</a> , <a href="#">Ort Global OmrådeVI</a> , <a href="#">Ort Ort OmrådeVI</a> , <a href="#">Ort Ort LandskapVI</a>
Pers	22 fält definierade, 92026 poster	<a href="#">Pers Pers Dublett</a> , <a href="#">Pers</a>
AccPers	14 fält definierade, 427854 poster	<a href="#">AccPers Acc Insamlare</a> , <a href="#">AccPers</a>
Roll	2 fält definierade, 8 poster	<a href="#">Roll</a>
Lank	15 fält definierade, 58816 poster	<a href="#">Lank</a>
Posttyp	2 fält definierade, 5 poster	<a href="#">Posttyp</a>
Org	5 fält definierade, 5 poster	<a href="#">Org Acc</a> , <a href="#">Org Acc Inloggad</a> , <a href="#">Org Anv</a> , <a href="#">Org Anv VI</a>
Användare	32 fält definierade, 64 poster	<a href="#">Användare Anv Hämta TillGlob</a> , <a href="#">Användare AccÄndradAv</a> , <a href="#">Användare AccSkapadAv</a> , <a href="#">Användare</a> , <a href="#">Användare Anv HämtaFrånGlob</a> , <a href="#">Användare Anv KontoCheck</a> , <a href="#">Användare AnvVisa</a>
AccKod	12 fält definierade, 105901 poster	<a href="#">AccKod KodNivå3</a> , <a href="#">AccKod KodNivå2</a> , <a href="#">AccKod KodNivå1</a> , <a href="#">AccKod</a>

KodNivå1	5 fält definierade, 23 poster	<a href="#"><i>KodNivå1</i></a> , <a href="#"><i>KodNivå1 AccKod</i></a>
KodNivå2	8 fält definierade, 114 poster	<a href="#"><i>KodNivå2</i></a> , <a href="#"><i>KodNivå2 AccKod</i></a> , <a href="#"><i>KodNivå2 KodNivå1_Vl</i></a>
KodNivå3	10 fält definierade, 43 poster	<a href="#"><i>KodNivå3</i></a> , <a href="#"><i>KodNivå3 AccKod</i></a> , <a href="#"><i>KodNivå3 KodNivå2_Vl</i></a>
FormGrund	2 fält definierade, 6 poster	<a href="#"><i>FormGrund</i></a>
Typ	3 fält definierade, 36 poster	<a href="#"><i>Typ</i></a>
FormDetalj	2 fält definierade, 23 poster	<a href="#"><i>FormDetalj</i></a>
PersDAUM	24 fält definierade, 24892 poster	<a href="#"><i>PersDAUM</i></a>
Socken	16 fält definierade, 2849 poster	<a href="#"><i>Socken Län38_Vl</i></a> , <a href="#"><i>Socken AccOrt_Sock</i></a> , <a href="#"><i>Socken Församl_Förs</i></a> , <a href="#"><i>Socken Ort</i></a> , <a href="#"><i>Socken Härad</i></a>
Härad	8 fält definierade, 322 poster	<a href="#"><i>Härad AccOrt_Härad</i></a> , <a href="#"><i>Härad Landsk</i></a> , <a href="#"><i>Härad Socken_Sock</i></a>
Landsk	7 fält definierade, 27 poster	<a href="#"><i>Landsk AccOrt_Landsk</i></a> , <a href="#"><i>Landsk Härad_Härad</i></a> , <a href="#"><i>Landsk Härad_Sock</i></a>
Län	7 fält definierade, 22 poster	<a href="#"><i>Län Kommun_Förs</i></a> , <a href="#"><i>Län Kommun_Sock</i></a> , <a href="#"><i>Län AccOrt_Län</i></a> , <a href="#"><i>Län Härad_Kom</i></a>
Kommun	8 fält definierade, 291 poster	<a href="#"><i>Kommun Förs_Förs</i></a> , <a href="#"><i>Kommun AccOrt_Kom</i></a> , <a href="#"><i>Kommun Län</i></a> , <a href="#"><i>Kommun Församl_Sock</i></a>
Län38	8 fält definierade, 25 poster	<a href="#"><i>Län38 AccOrt</i></a> , <a href="#"><i>Län38 Socken_Sock</i></a>
Län88	7 fält definierade, 25 poster	<a href="#"><i>Län88</i></a>
Församl	9 fält definierade, 1493 poster	<a href="#"><i>Församl Kommun</i></a> , <a href="#"><i>Församl AccOrt_Förs</i></a> , <a href="#"><i>Församl Socken_Sock</i></a>
Disciplin	3 fält definierade, 9 poster	<a href="#"><i>Typ 2</i></a>
Land	3 fält definierade, 257 poster	<a href="#"><i>Land 2</i></a> , <a href="#"><i>Land AccOrt</i></a>
Realkatalog	23 fält definierade, 123830 poster	<a href="#"><i>Realkatalog</i></a> , <a href="#"><i>Signum_Rubrik4Glob</i></a> , <a href="#"><i>Realkatalog</i></a> , <a href="#"><i>Signum_Rubrik3Glob</i></a> , <a href="#"><i>Realkatalog</i></a>
Signum	19 fält definierade, 5139 poster	<a href="#"><i>Signum</i></a> , <a href="#"><i>Signum Rubrik3_Global</i></a> , <a href="#"><i>Signum</i></a>

		<a href="#"><i>Rubrik4 Global, Signum Real VI Drill</i></a>
Rubrik_0Huvud	3 fält definierade, 22 poster	<a href="#"><i>Huvudrubrik Global, Huvudrubrik</i></a>
Rubrik0	7 fält definierade, 132 poster	<a href="#"><i>Rubrik0 Huvudrubrik Global, Rubrik0 Real VI Drill, Rubrik0</i></a>
Rubrik1	8 fält definierade, 867 poster	<a href="#"><i>Rubrik1 Signum VI, Rubrik1 Rubrik0 Global, Rubrik1 Real VI Drill, Rubrik1</i></a>
Rubrik2	6 fält definierade, 581 poster	<a href="#"><i>Rubrik2, Rubrik2 Rubrik1 Global, Rubrik2 Real VI Drill</i></a>
Rubrik3	6 fält definierade, 250 poster	<a href="#"><i>Rubrik3, Rubrik3 Rubrik2 Global, Rubrik3 Real VI Drill</i></a>
Rubrik4	6 fält definierade, 21 poster	<a href="#"><i>Rubrik4 Rubrik3 Global, Rubrik4, Rubrik4 Real VI Drill</i></a>

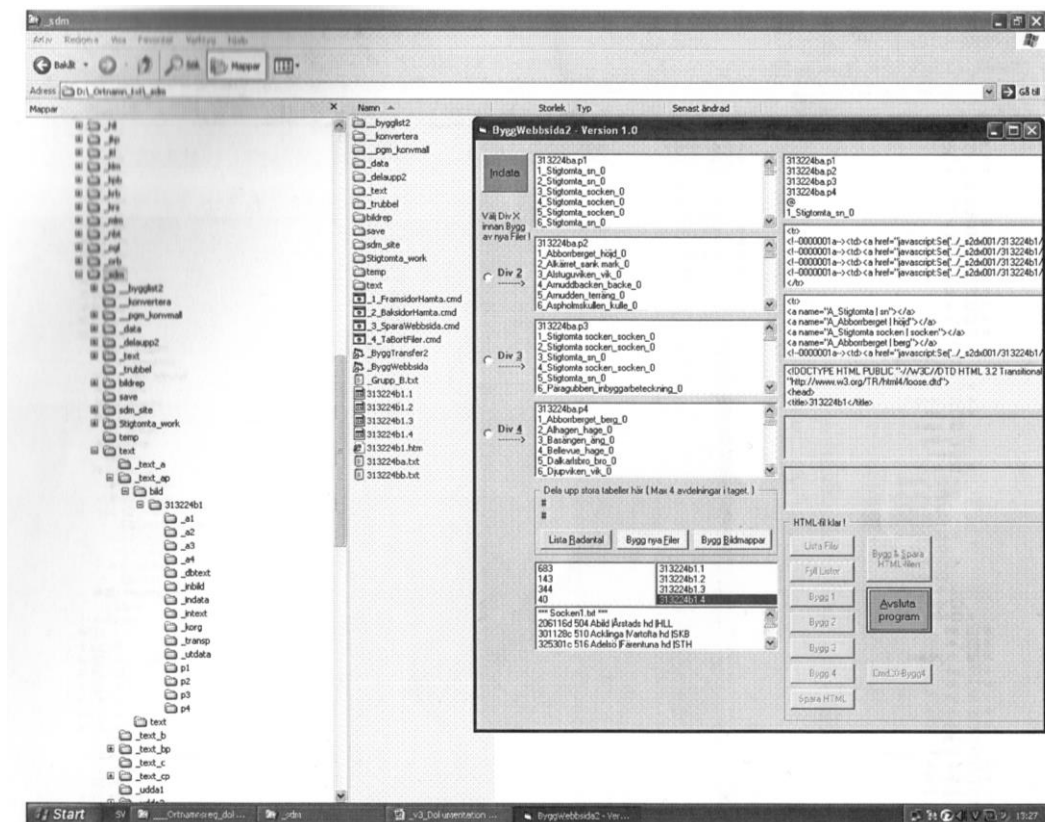
### 3.2 Mjukvara

ISO/IEC JTC1 har överhuvudtaget en konservativ och genomtänkt mjukvarupolitik, uppdateringarna är kanske den svaga punkten i politiken just nu, men framtiden bör ändra på saken. MySQL, PHP, XML, PDF, PNG osv. är standarder och långvariga mjukvaruval med få problem. Man kan kanske även säga att Microsoft Word (och Office) som ett val för dokumentskrivandet är bra, för Word (och Office) är de facto standard(er), och människor i allmänhet är så pass vana med de här programmen att de sällan tänker på alternativen.

Databassystemet som Ortnamnsregistret utnyttjar har följande egenskaper som man ser viktiga i det här sammanhanget. OS är unix. MySQL server -versionen är 5.0.91, databasmotorn är MyISAM, teckenkodningen är latin1. MySQL latin1 är identisk med Windows cp1252 (Oracle Corporation and/or its affiliates 2014b). PHP-version är 5.2.17. Sedan har vi PHP-biblioteken. Det enda biblioteket som rör oss här är mysql.

Forskarna använder och har använt olika Windows-versioner och olika browser med ortnamnsdatabaserna. Det har varit få problem med olika operativsystem och browserversioner.

Inmatningssystemet har 13 steg som innehåller till exempel att kopiera filer, konvertera filer, köra hjälpprogram (Sigvardsson 2014, 9-25). I figur 7 ser man steget 10.



Figur 7. Ortnamnsregistrets inmatningssystem (Sigvardsson 2014, 19)

Forskarna och lekmännen söker i Ortnamnsregistret med hjälp av ett PHP-program. Man kan se att sökprogrammet erbjuder omfattande sökmöjligheter, programmet är väl planerat och implementerat, det centrala är att man har goda sökmöjligheter gällande orter. Programmet söker ortnamnsinformation i en MySQL-databas. Programmets GUI ses här (Namnarkivet i Uppsala 2014 och figur 8):

**Sök i hela Ortnamnsregistret**

[Till registrets startsida](#) [Till information](#) [Till Skånskt ortnamnsdatabas »](#)

**Sök på**

Ortnamn % (uteslut i Ortnamn)

Lokal % (uteslut i Lokal)

Socken %

Härad %

Län %

Landskap %

**Välj kategori**

Samtliga	Äldre namnformer	Uppteckningar
<input checked="" type="radio"/> Bebyggelse- och naturnamn	<input type="radio"/> Bebyggelsenamn	<input type="radio"/> Naturnamn
	<input type="radio"/> Bebyggelsenamn	<input type="radio"/> Naturnamn

**Välj språk**

<input checked="" type="radio"/> Alla	<input type="radio"/> Svenska	<input type="radio"/> Finska	<input type="radio"/> Samiska
---------------------------------------	-------------------------------	------------------------------	-------------------------------

☐ Visa endast tolkningar

**Sök**

**Sök tips:** med % menas vilket antal tecken som helst och med understreck( \_ ) ett enskilt tecken utelämnat. Önskas uteslutning av ett/flera tecken använd uteslut-textrutorna.

[Skriv ut](#)

Figur 8. Ortnamnsregistrets sökprogram

Det finns problem med sökfunktionernas snabbhet därför att den varierar. Det är kanske svårt att säga varför, för samma sökning kan gå relativt snabbt eller mycket långsamt.

### 3.2.1 Uppdateringspolitik

IT-avdelningen har ingen uppdateringspolitik för tillfället. Orsaken är att till hösten ska många saker förändras. Några saker som skildras här kommer att inte vara något problem, men nu har man tid att formulera en ny uppdateringspolitik.

Eftersom ISOF inte behöver serverfarmar och knappast stort lockar den allmänna publiken, finns det inte ett stort behov av att akta väldigt paranoida med serverna och mjukvaran.

## 4 TANKAR OM FÖRÄNDRINGAR OCH FÖRSLAG

### 4.1 Databasförändringar

Det är svårt att tänka på vilka kolumner man borde ha mer i Ortnamnsregistret. Man kan nog tänka på några, men att lägga till de där nya kolumnerna i databasen betyder att man behöver mera personal, tid och pengar, kanske rum och några småsaker, och det är inte sagt att det lönar sig. Här är dock några exemplar som man kan tänka på som (långtids)mål.

Man kan systematiskt kartlägga de dialektala formerna. Man kan gå igenom arkivet och mata in fonologiska eller fonetiska former av ortnamn med tydliga luckor. Det här skulle vara ett enormt projekt, dock under ett långt tidsspann är det genomförbart. Man skulle säkert lära sig något nytt det är oundvikligt men sannolikt kommer projektet inte att kunna genomföras av resursskäl. Det är nog möjligt att kartlägga något av intresse man kan ta ett mindre ambitiöst mål till exempel låt oss säga omkring och i de områdena där älvdalskan talas (andra dialekter och språk med ett begränsat geografiskt läge kommer i fråga, såsom gutniskan och sydsamiskan).

Ett mål kunde vara att ha en runortnamnsdatabas. Det finns också nytta av eller åtminstone behov för en runortnamnsdatabas. Låt oss låna kursbeskrivningen för en kurs "Ortnamn, runor, dialekter. Runor" (Wise 2010):

*Denna kurs behandlar dels runorna som skriftsystem, dels själva inskrifterna, som till stor del studeras i ett östgötskt perspektiv. Den handlar också om hur östgötarnas liv och leverna under vikingatiden avspeglas i runinskrifterna och vad dessas placering kan berätta om det östgötska kulturlandskapets framväxt.*

Bra är att vi har omfattande källor (Peterson 2002). Det är också bra att databasen egentligen skulle bli ytterst liten. Peterson (Peterson 2002, 288) listar bara 157 ortnamn i nordiska runor (i vikingatiden).

Man kan lägga till en kolumn med en ytlig morfologisk analys gjord med datorlingvistiska program. Även om analysen i några fall kommer att bli värdelös,

kommer den också att hitta riktiga mål: ortnamnen är ju mycket ofta normaliserade i databaserna – Ortnamnsregistret bör egentligen innehålla bara normaliserade ortnamn – och därför kan vi också ofta göra en riktig ytlig morfologisk analys. Naturligtvis kan en sådan analys bara vara en början på forskningen forskaren ska tolka det som man har funnit. Men också här kan IT hjälpa en vidare; några möjligheter är anpassning till dialekten med dialektala regler och att begränsa sökningen till de relevanta orterna. Man borde ändå minnas att

- datan är dock inte alltid enhetlig
- resultatet berättar knappast mycket utan analys
- analysen är någon gång kanske ytligt helt korrekt, men det är bara en halv historia (såsom i fallet *Vedyxa*).

#### 4.1.1 Beläggsår

En kolumn till vill forskarna ha i Ortnamnsregistret, nämligen beläggsåret. Det kan kännas att det inte betyder mycket när ett ortnamn är belagt, men det finns mycket att tänka på: transkription, kriterier för belägget, även personlig information.

Egentligen finns den här kolumnen redan i Ortnamnsregistret, eller rentav två kolumner: beläggsåret och vidare beläggsår, men de här kolumnerna är ännu tomma.

I princip finns det många teknologiska möjligheter att tillfoga beläggsåret. Det har länge varit möjligt. Alla i allmänhet vet om OCR (*optisk teckenläsning*), kanske inte med det namnet, men man ser till exempel dessa kontrasigneringsboxar överallt på Internet. Handskriven text är inte heller något hinder. Teknologin bakom OCR är mogen och mycket bra, detta sedan årtionden. De inskannade sedeskorten är perfekta för OCR. Att identifiera bara åren hjälper också mycket. Om kvaliteten inte är bra, kan man också lära OCR med data. Det ända ”men” i detta i princip är att det kan ta sin tid att fylla beläggsåren i den nuvarande servermiljön (jämför omvandlingen från PDF till PNG som äger rum). Detta är en orsak till att göra något åt servermiljön.

Alla borde egentligen experimentera med OCR för att se hur utmärkt identifieringen lyckas. En liten personlig historia följer för att illustrera saken. För några år sedan ville jag själv ta reda på hur bra OCR egentligen fungerar. Jag tog den ungerska klassikerromanen *Egri csillagok* och skannade in den första sidan med min konsumentskanner. Sedan matade jag den inskannade filen in i ett webb-OCR-program. Resultatet var perfekt. Det gick verkligen inte att skilja från originalet (inga fel, fonterna ser ut som i originalet, textformateringen stämmer med originalet).

#### 4.1.2 Etymologisk forskning

Man kan kanske tillfoga databasen en kolumn som har med etymologi att göra. Det vore ändå bättre att göra en egen etymologisk databas av ortnamn. Man kan då fullt ut använda det som gjorts tidigare inom det datorlingvistiska området. Det finns en etymologisk XML-databas, Uralothek, som gjorts i Koblenz, Tyskland. Databasen är ett resultat av forskning av uralisk språkhistoria över decennier och sedan en omstrukturering från bokform till databasformat. Strukturen tar hand om det som behövs för att åskådliggöra etymologier. Även om man har lite begränsade behov i Ortnamnsregistret, är Uralothek en utmärkt modell för etymologidatabasstrukturen, och också POC. Databasen ser man i figur 9. (Universitt Koblenz-Landau 2014.)

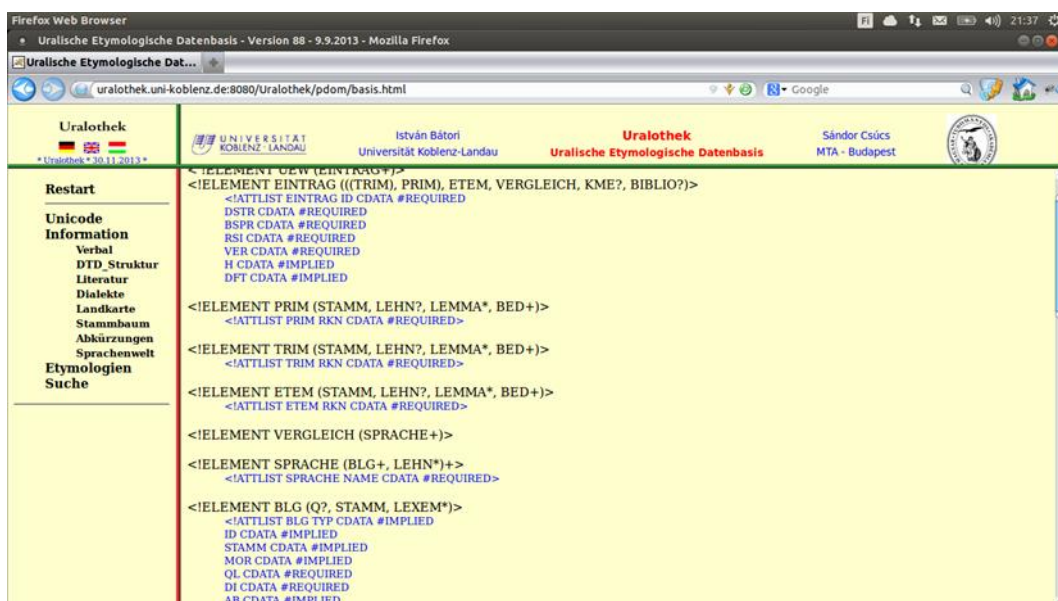
The screenshot shows a Firefox browser window with the URL [uralothek.uni-koblenz.de:8080/Uralothek/pdom/basis.html](http://uralothek.uni-koblenz.de:8080/Uralothek/pdom/basis.html). The page title is "Uralische Etymologische Datenbasis - Version 88 - 9.9.2013 - Mozilla Firefox". The page content includes a header with logos for the University of Koblenz-Landau and the Uralothek project, and a main table titled "Aufzhlung der rekonstruierten Formen der uralischen Etymologien".

Nr.	Rek.-Ebene	Stamm	Bedeutung	Anzahl der belegten Sprachen	Anzahl der Belege	Perm-Id.Nr.
1	FP fu	<i>ača</i>	1: (Bach)wiese, 2: Tal	4	6	
2	fu	<i>ačka</i>	1: weiß	5	10	
3	FU	<i>ačs, (očs)</i>	1: ein mit etwas bedeckter, zum Liegen (Schlafen) geeigneter Platz, 2: durch Bedecken einen Platz zum Liegen / Schlafen bereiten, 3: betten	4	11	
4	FU	<i>aja</i>	1: treiben, 2: jagen	6	17	
5	FW fu	<i>aļpe</i>	1: Schläfe	5	8	
6	FU	<i>akta</i>	1: aufhängen, 2: stecken, 3: stellen Falle, Netz	8	21	
7	Ur	<i>ala</i>	1: Raum unter etw., 2: Unter-, 3: das Untere	15	37	
8	FU ur	<i>alka</i>	1: vorderes od. hinteres Ende, 2: Anfang, 3: anfangen, 4: beginnen	8	26	
9	FU	<i>als</i>	1: Zaubersprüche hersagen, 2: verzaubern	4	6	
10	FU	<i>ama</i>	1: schöpfen	5	11	
11	FU	<i>amta</i>	1: geben	7	13	
12	Ur	<i>ams</i>	1: sitzen	7	18	
13	ur	<i>an3, on3</i>	1: groß, 2: viel	4	12	
14	FU ur	<i>an3, an3-pp3</i>	1: Schwiegermutter	10	17	

Figur 9. Uralothek



Det finns information om databasen på den ovannämnda hemsidan, till exempel är DTD-strukturen relevant här. Så ser en del av den ut (figur 10):



Figur 10. DTD för etymologier i Uralothek

Den ursprungliga källan (egentligen tre böcker) för Uralothek är Uralisches etymologisches Wörterbuch (Rédei 1988–1991). En exempletymologi ur boken (Rédei 1988, 119) ser man i figur 11.

#### **kala** 'Fisch' U

Finn. *kala* 'Fisch'; est. *kala* | lapp. N *guolle -l-*, L *kuollē*, K (328) T *ki̯lle*, Kld. *kūll*, Not. *kuo̯ll* | mord. E M *kal* | tscher. KB U B *kol* | ostj. (OL 281) V *kul*, DN *χur*, O *χul* (Vj. *nārək kul* usw. 'roher Fisch' > syrj. I Ob *nārχul*) | wog. (WV 115) TJ *kōl*, KU So. *χūl*, P *kul* | ung. *hal* | sam. jur. (165) O *χāl'e*, Nj. *kāre*; jen. Ch. *kađe*, (DONN.—JOKI: JSFOu. 58/1: 13) Ch. *karē*, K *kare* B *kare*; twg. *kole*, (DONN.—JOKI: a.a.O. 16) *kuale*, *kuälle*; selk. (DONN. Mskr.) TaM *keḷ*, *qeḷ*, Ty. *keḷ*, (MSFOu. 49: 32) *qeḷi*, Ke. *ḷuel*, (LEHT.: MSFOu. 122: 320) Tur. *qēli*: kam. *kola*; koi b. (KLAPR., mitg. DONN.: MSFOu. 64: 32, 41) *kola*; mot. (ebd.) *kele*; karag. (PALL., mitg. DONN.: MSFOu. 64: 17) *kalē*; taig. (KLAPR., mitg. DONN.: MSFOu. 64: 41) *kallā*.

Vgl. ? alt.: ma.-tung. \**k'olo* ~ \**k'olto*: tung. *ollo*, *oldo*, *olro*, lam. *olra* 'Fisch', orok. *χolto*, ude *oloxo*, gold. *χolto(n)* 'gekochter Fisch'; ieur.: lat. *squalus* 'ein größerer Meerfisch' (< \*(s)qalo).

SAJN., Dem. 113; GYARM., Aff. 206, 216, 270; MUSz. 77; VglWb. 202; HALÁSZ: NyK 23: 27; WIKLUND: MO 1: 57; Beitr. 38; NyH7; LEHTISALO: FUF 21: 14; SzófSz.; RÄSÄNEN: Vir. 1947: 166; FUV; SKES; RÉDEI: NyK 66: 7; MSzFgrE; TESz.

Figur 11. En etymologi i bokform (Rédei 1988, 119)

## 4.2 Hårdvara

En egen server för IT-avdelningen vore kanske inte en dålig idé, eftersom då kunde man förbereda sig för att göra uppdateringar i en miljö som speglar det reala systemet. Det går kanske med en vanlig dator som server, men om man i den produktions-servern har till exempel ett lite sällsynt val av operativsystem, så kan det vara svårt att emulera saker.

Man har möjligheter att diagnosticera server-problem, speciellt prestanda, åtminstone om man kan arbeta i servern med root-rättigheterna. Naturligtvis kan man börja med till exempel *top* men flera program har testuppsättningar, såsom MySQL har MySQL Test Programs (Oracle Corporation and/or its affiliates 2014c). Men som sagt om man har en begränsad IT-miljö, så är kanske också diagnosmöjligheterna och korrigeringsmöjligheterna begränsade.

### 4.2.1 Uppdateringspolitik

Politiken över uppdateringarna har också mycket att göra med hårdvaran. Man måste ju först och främst ha en ordentlig möjlighet att uppdatera. Då borde man tänka på vilka prioriteringar man har: är det bättre att ha ett fungerande men åldrande system än att uppdatera systemet och kanske korrigera många och okända aspekter av den mjukvarumiljön?

## 4.3 Mjukvara

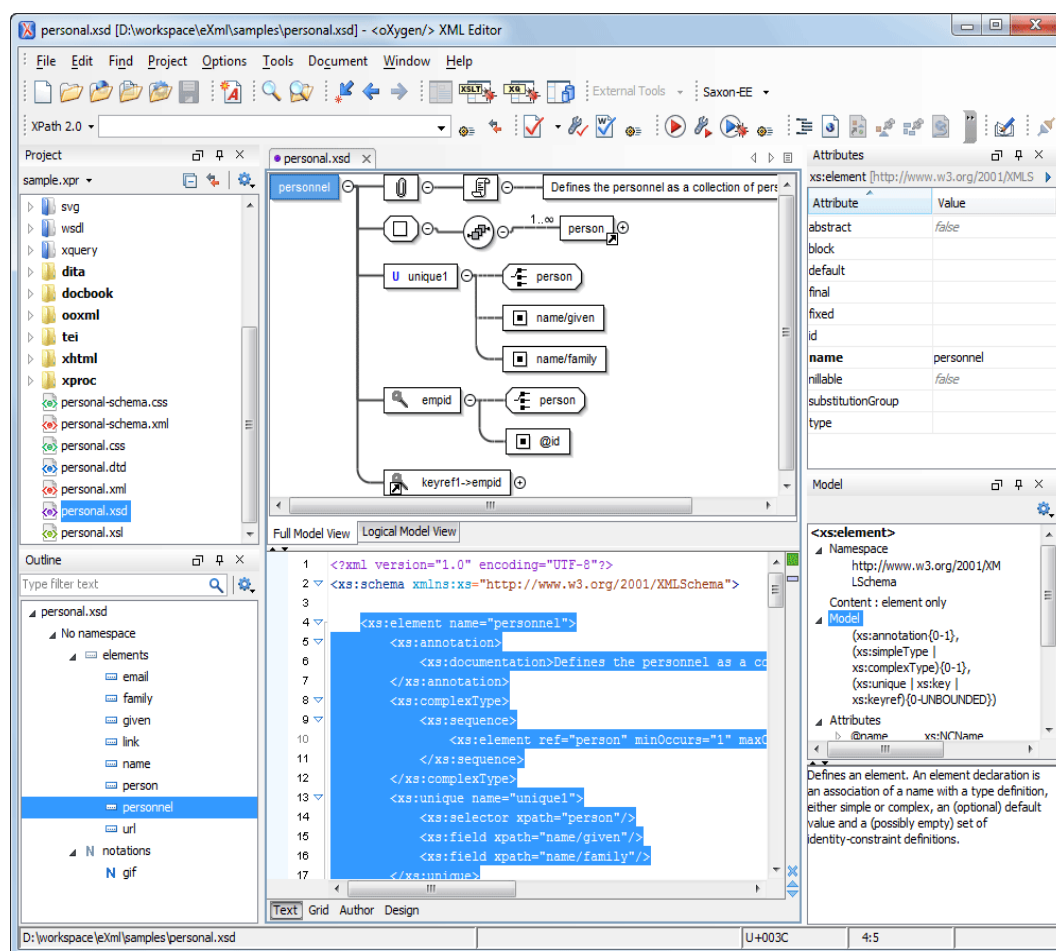
### 4.3.1 Inmatningsprogram

Man har ny data som väntar på att matas in i databasen. Det behövs ett nytt inmatningsprogram, eller åtminstone har den nuvarande inmatningsmetoden kritiserats. Det finns en början på ett sådant program men inte något som är fritt användbart för allt som behövs just nu.

Ny data som behövs för att mata in är framför allt sydsvenska ortnamn. Datan i sig själv är likadan som i det nuvarande Ortnamnsregistret, alltså databasstrukturen ska vara densamma som den nu är.

Man kan alltså antingen använda och kanske försöka förbättra programmet som finns, eller bygga ett nytt program till exempel med hjälp av PHP-programmet som används för att söka i Ortnamnsregistret (i princip är sökfunktionerna och inmatningsfunktionerna lätta att byta i koden när det handlar om SQL).

Om man gör en XML-databas, kan man fråga om ett inmatningsprogram behövs överhuvudtaget. Man kan ju använda vad som helst för att mata in XML-enheter. Visst är det bekvämare att använda ett program som man har anpassat till det här ändamålet, men kanske genererade XML-filer och en WYSIWYG-editor är tillräckligt goda. Man kan välja från många olika alternativ, till exempel i figur 12 syns oXygen.



Figur 12. En WYSIWYG-XML-editor oXygen

XML-enhetsfiler löser de behov och bekymmer som Vikstrand har med insamlingsgränssnitt, redigeringsgränssnitt (för ny data; för befintligt material är

saken lite mer komplicerad) och FileMaker (Vikstrand 2013). De kan lätt vara projektspecifika, de kan integreras till vidare användning när de är färdiga och de kan förändras fritt så ofta som man vill. Om FileMaker kan användas som pålitlig XML-editor, då finns det i princip ingenting emot att använda FileMaker med ortnamnssamlingarna. FileMaker är bara lite dyr (licensvis) som en XML-editor. Saken med nytt material är bara det att det antingen måste innehålla material strukturellt precis som Ortnamnsregistret innehåller eller så bör Ortnamnsregistret ändras d.v.s. göras bredare så att det extra materialet bara kan bifogas (då båda strukturellt blir precis desamma). Om dessa villkor inte är uppfyllda, uppstår dataförlust.

Man måste tänka på redigeringen av befintligt material. Hur kommer man på det man ska redigera? Om man besvarar frågan och analyserar situationen med Ortnamnsregistret, är det sannolikt så att inget eget redigeringsgränssnitt alls behövs. Om det gäller isolerade och sällsynta fall så borde man ta upp identifierande information så att IT-avdelningen ska kunna korrigera saken. För att överhuvudtaget korrigera (enkla) fel skulle man som en enkel förstastegslösning helst göra scripter anpassade för jobbet.

Det finns ett problem som kanske inte kan lösas helt och hållet. Vikstrand säger (Vikstrand 2013, 3):

*Samlingar som kommer in idag har ofta formen av excellark.  
Också den typen av material är det viktigt att vi enkelt kan föra  
in i Ortnamnsregistret.*

Det finns för lite information för att kunna ge ett tillfredställande svar här. Om de inkommande samlingarna inte är strukturerade eller strukturen varierar, så bör man antingen göra inmatning för hand eller använda någon sorts "rule-based" eller sannolikhetsbaserad algoritm att ta reda på vad som är vad därinne. Programmet Marshal, som man på IT-avdelningen möjligen funderar på att skaffa, gör någonting sådant. Om samlingarna är små och kommer sällan så matas de troligen in för hand. Om det finns en enhetlig struktur eller om de är indelade i få delar som har få enhetliga strukturer så går i princip inmatningen automatiskt genom att använda ett konversionsscript eller ska gå efter att man gjort ett konversionsscript.

Vikstrand pratar om Användargränssnittet (Vikstrand 2013). Han föreslår en klickbar karta över Sverige (och varför inte Åland osv. också) för att söka eller mata in material. Vikstrand säger (Vikstrand 2013, 2):

*En fördel med ett sådant system är att det är väldigt flexibelt och att nya material lätt kan tillfogas.*

Man antar att nya material redan innehåller väsentlig ortsinformation. Om det inte är så är någonting på tok och ett sådant material bör i alla fall inte inmatas. Således finns det i princip inget behov att mata in ortsinformation på nytt genom en klickbar karta eller på något annat sätt. Det finns andra möjligheter här. Konversionsskript kan ofta göra jobbet eller man kan generera XML-enhetsfiler där den väsentliga (orts)informationen är tillfogad.

#### 4.3.2 Sökprogram

Visst är det möjligt att förändra det nuvarande sökprogrammet. Det är svårt att veta i vilken riktning man ska gå med sökfunktionerna, det är inte klart alls. Kanske finns det inga konkreta idéer? Det är inte omöjligt att någon forskare kommer att kräva något som kan lösas med en ny sökfunktion men troligen handlar det då om ett fall på tusen. I alla fall kan man säga att om man vill ha mera sökmöjligheter så är det förnuftigt att använda MySQLs sökmöjligheter.

Naturligtvis kommer man att ändra sökprogrammet om man kommer att ha nya kolumner i databasen. Det är dock mer underhåll än att programmera någonting nytt.

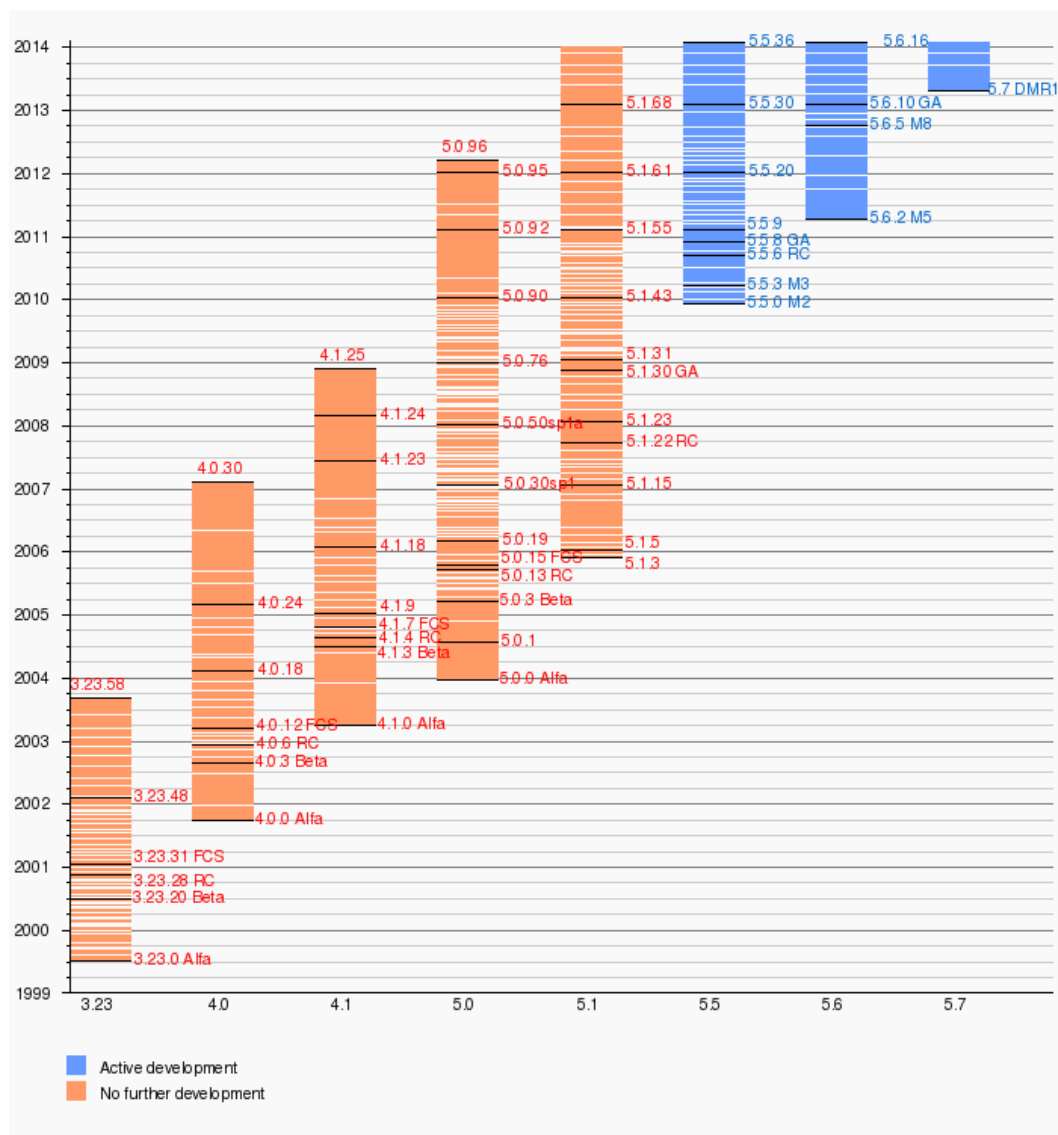
Man har nu en ny POC-version av sökprogrammet som är ytligt såsom förr. Programmet fungerar med nya versioner av PHP och MySQL. Alltså ISOF är färdigt att uppdatera servermiljön.

Det som Vikstrand (Vikstrand 2013) gärna skulle vilja ha i sökprogrammet (Användargränssnittet) är i sig ingenting nytt. Det kan man göra med det nuvarande programmet, men är inte alls lätt. Det är visserligen en förbättring som erbjuder det som är av nytta för forskarna. Att lätt välja ort som man vill söka är en bra idé. Sockennivån (1951) är också den nivå som önskas. Man måste

sannolikt omprogrammera mycket i sökprogrammet. Det finns troligen många möjligheter med en klickbar karta, en som man kan tänka på är ImageMapster (Treworky 2014). Man har nu ett sökprogram som fungerar och detta förslag kan också förverkligas. Ett viktigt sidoprojekt med en integrerad karta-GUI är en bra idé.

#### 4.3.3 MySQL

Två relativt nya saker gör MySQL bättre och man borde uppdatera gamla versioner. I version 5.6.4 och senare kan man söka text fritt med UTF-8 med databasmotorn InnoDB (detta var en av få fördelar med databasmotorn MyISAM). Ganska nytt är också att den förinställda databasmotorn nuförtiden igen är InnoDB (från version 5.5 och uppåt). Det finns också andra orsaker att uppdatera (Bradford 2011 och Schwartz 2011). MySQL har haft en lång supporttid (Wikipedia 2014a) som man ser i figur 13. MySQL-versionen som Ortnamnsregistret använder har nått supporttidens slut för två år sedan. För datasäkerhetens skull bör man uppdatera så snabbt man kan.



Figur 13. Supporttiden för MySQL-versioner (Wikipedia 2014a)

Det är värt att notera att man nästan alltid borde välja InnoDB som databasmotor.

Det är en allmän professionell åsikt (Schwartz, Zaitsev och Tkachenko 2012, 15):

*InnoDB is the default transactional storage engine for MySQL and the most important and broadly useful engine overall. It was designed for processing many short-lived transactions that usually complete rather than being rolled back. Its performance and automatic crash recovery make it popular for nontransactional storage needs, too. You should use InnoDB for your tables unless you have a compelling need to use a different engine.*

Det är kanske värt att konvertera FileMaker-filer till MySQL-format. Man kan också använda program som gör användningen av (alla) databaser, dokument etc

ytligt enhetlig. Marshal, som är gjort av en uppsaliensisk firma, kan göra sådana saker (R2B Software 2014). Det är ett alternativ värt att tänka på, såsom man gör nu på IT-avdelningen.

Här bör det även berättas om POC-testfall. MySQL har sina egna testprogram (Oracle Corporation and/or its affiliates 2014c) som man troligen alltid har installerad med MySQL. I detta fall hade man en normal Ubuntu 13.10 installerad i en laptop. MySQL var i version 5.5.35 som hade anpassats till Ubuntu. Meningen var bara att gå igenom alla tester med scriptet `mysql-test-run.pl`. Scriptet slutade fungera alldeles i början. Efter att ha studerat saken hittades en lösning (Launchpad 2012). Här ser vi att det är bättre att ha kritiska program installerade så ursprungliga som möjligt, distroanpassade versioner kan bära med sig obekväma överraskningar. Och detta är också något som man borde tänka på i samband med uppdateringspolitiken.

#### 4.3.4 PHP

Man borde egentligen använda en senare version av PHP, för versionen i servern har uppnått supporttidens slut (End-of-life) för tre år sedan och har inget officiellt stöd längre (The PHP Group 2014b). En ny PHP-version (minor version) släpps en gång om året och supporttiden är tre år (Wikipedia 2014b), se tabell 2. För datasäkerhetens skull bör man uppdatera så snabbt man kan. Det finns snabbhetstester och minnestester över PHP-versioner (Lourdass 2013). Och visst finns det skillnader där, men skillnaderna är minimala och man måste anpassa resultaten till den verkliga världen.

Tabell 2. Supporttiden för PHP-versioner (Wikipedia 2014b)

Color	Meaning	Development
Red	Old release	No development
Yellow	Stable release	Security fixes
Green	Stable release	Bug and security fixes
Blue	Future release	New features



Version	Release date	Supported until
1.0	8 June 1995	
2.0	1 November 1997	
3.0	6 June 1998	20 October 2000
4.0	22 May 2000	23 January 2001
4.1	10 December 2001	12 March 2002
4.2	22 April 2002	6 September 2002
4.3	27 December 2002	31 March 2005
4.4	11 July 2005	7 August 2008
5.0	13 July 2004	5 September 2005
5.1	24 November 2005	24 August 2006
5.2	2 November 2006	6 January 2011
5.3	30 June 2009	July 2014
5.4	1 March 2012	1 March 2015
5.5	20 June 2013	20 June 2016
5.6	No date set	3 years after release

PHP-biblioteket `mysql` är deprecated nuförtiden (det rekommenderas inte), och man har länge propagerat för `mysqli` (och också för `PDO`). Det är mestadels ganska lätt att gå över till `mysqli` (SitePoint 2013), att ha `mysqli` i stället för `mysql` räcker ofta. `mysqli` är också snabbare än `mysql` (ioMeWeekly 2013).

Det finns andra små ändringar i PHP. `import_request_variables()` är inte längre där (från PHP-version 5.4.0), saken resulterar i att sökprogrammet inte fungerar längre med nya versioner av PHP (The PHP Group 2014a). Man kan använda `extract()` (The PHP Group 2014a) i stället. En annan sak är att `$PHP_SELF` borde vara `$_SERVER['PHP_SELF']`.

#### 4.3.5 Testfall

Testfall är viktiga för databasintegriteten kort sagt. Man vill ha en databas som har den riktiga informationen därinne. Testfallen bör alltså först och främst producera de förväntade resultaten. Man bör också sträva efter omfattande databaser. Omorganisation och dataförändringar borde bara påverka datan som är vald för förändringar. Inom dessa ramar är det likgiltigt vad testfallen innehåller. Med

ortnamnen kan man egentligen ta exemplariska testfall, för databaserna är väldigt stora.

Testfallen kan man också använda för att kolla upp hur effektiva sökfunktionerna är. Snabbhet och kodens lämplighet kan testas med nästan likadana funktioner.

Låt oss säga att databasen har tre kolumner (x, y och z) i en tabell, så kan vi se hur annorlunda "SELECT \*" och "SELECT x, y, z" är osv. IT-miljöns brister kan också förekomma med omfattande tester, till exempel kan servern priorisera andra saker än databasanvändning.

Låt oss enumerera några exempel eller kvasi-specificeringar (innehållet är ju i sig likgiltigt) av testfall. 1) Alla sjöar i registret med namnet Långsjö. 2) Alla socknar med strand i Väner. 3) Alla Drottninggator i Dalarna. 4) Alla fjäll i Stockholm, meningen är att det inte finns några. 5) Något som är "nästan alla": alla ortnamn utom fjällen.

#### 4.4 Andra förändringar

Här beskrivs några mindre förändringar som inte är så viktiga. Några kommer att bli viktiga om man kommer att implementera andra förändringar. Om man, till exempel, kommer att mata in dialektal information då behöver man egentligen en UTF-8-teckenkodning för det fonetiska alfabetet. Man måste också då tänka på fonterna. Teckenkodning är också viktig av andra skäl, kanske en viktigare sak än man kan tro.

##### 4.4.1 Fonter

Man borde tänka på, om det finns tecken eller delar av tecken som man inte kan se tydligt i texten. En annan font kan hjälpa. Den generella editorn gedit i (den amerikansk-engelska varianten av) Ubuntu 13.10 (och före), till exempel, har en standard font där <å> och <ä> ser ganska lika ut.

Om man kommer att skriva något i databasen utanför de svenska bokstäverna kommer fonterna att vara något att verkligen tänka på. Lingvistiska och fonetiska aspekter är inte sådana aspekter som man ofta tänker på när man planerar fonter,

därför är normala fonter inte alltid lämpliga för ISOFs forskare. För att se hur det kan se ut, även idag, se figur 14 där tundranentsiskan skrivs med det uraliska fonetiska alfabetet (Lehtisalo 1947, 65), diakritiska tecken är av speciell betydelse. Trots allt är problem med fonterna få med det nuvarande databasläget.

## IV. Erzählungen und Märchen von Tieren.

25.

š̌ ō š̌ a β̌ a ǐ ľ a χ̌ ň ā ǩ ǩ u.

lēm̌morō·lťp̌ ňaŋ̌ni, ňiš̌'eďp̌ β̌ēš̌oǩko, ňēb̌'eďp̌ puχ̌ū·ťš̌e. ηoṑṑoṑi  
 ňūď'i''<sup>2</sup>. p̌iš̌'im̌-p̌iřip̌ťš̌e taŋ̌ňā. lēm̌morō·lťp̌ ňaŋ̌ni mā'' ňiβ̌''<sup>3</sup>: »p̌i-  
 řip̌ťš̌e taŋ̌ňā, ňēdaŋ̌i'' m̌ēη̌Ǧūβ̌.« mā: »ǰā ňālťp̌ ňōβ̌ŋ̌ňp̌ ǰādeřp̌-  
 p̌alťp̌ š̌aβ̌β̌o η̌āēš̌'eti, χ̌ām̌š̌aβ̌, ňēdaŋ̌i'' m̌ēη̌Ǧūβ̌.« ňiš̌'eďp̌ māŋ̌ňq̌m̌p̌:  
<sup>5</sup> »η̌āǐβ̌āř ǰoχ̌ōďā.« (ňiš̌'eďp̌ māŋ̌ňq̌m̌p̌:) »taŋ̌ňē m̌ēǰje!« »maŋ̌ χ̌ānťā·dm̌''<sup>4</sup>,  
 m̌ūη̌Ǧanť ňū·dm̌'' ňq̌mťaη̌ku''<sup>5</sup>. ňēdaŋ̌i'' m̌ēβ̌β̌āň'' ňiň χ̌ařβ̌ā''<sup>6</sup>.  
 ňiš̌'eďp̌ māŋ̌ňq̌m̌p̌: »ňēdaŋ̌m̌p̌ χ̌ařni'' m̌ēη̌Ǧū·dm̌''<sup>7</sup>.« lēm̌morōlťp̌  
 ňaŋ̌ni ǰāŋ̌ňq̌kaŋ̌je, p̌iš̌'im̌-p̌iřip̌ťš̌e maŋ̌ňēďp̌, š̌aβ̌β̌o p̌iřip̌ťš̌e  
 η̌āǐβ̌β̌i. lēm̌morōlťp̌ ňaŋ̌ni ňālť'' ťš̌u. ňē š̌āǐdořā, p̌iβ̌β̌om̌'' š̌āǐďǐbi.  
<sup>10</sup> p̌iš̌'im̌-p̌iřip̌ťš̌eη̌ǩēβ̌χ̌aŋ̌ňq̌η̌āmť''<sup>8</sup>. p̌iš̌'im̌-p̌iřip̌ťš̌e ǰuřkǩi''<sup>9</sup>,  
 p̌iš̌'im̌-p̌iřip̌ťš̌e χ̌āsoβ̌β̌aňľp̌ m̌ē š̌ēřiťā, š̌āim̌'' p̌iřē. η̌aβ̌β̌orňp̌.

25.

Märchen (zu singen).

Ein unverheirateter Mann (Junggeselle) ist der Spatz. Sein Vater  
 ist der Alte, seine Mutter die Alte. Sie haben einen Sohn. Es  
 ist ein Mäusefräulein. Junggeselle Spatz sagte: »Die Jungfrau ist  
 vorhanden, ich nehme sie zu meinem Weib.« Er sagte: »Wenn sie  
 durch die Tür ihres Erdzeltes geht, ist sie gut. Ich liebe sie. Ich  
 nehme sie zu meinem Weibe.« Sein Vater sagte: »Dein Kopf geht zu  
<sup>5</sup> Grunde. Mag sie dort sein!« »Ich gehe fort, ich höre nicht auf deine  
 Stimme, du willst für mich keine Frau nehmen.« Sein Vater sagte:  
 »Ich nehme selbst für dich ein Weib.« Der Spatz, der unverheiratete  
 Mann (Junggeselle) begab sich auf die Erde. Er sah das Mäusefräu-  
 lein, es war ein gutes Fräulein. Junggeselle Spatz ging hinein in das  
 Zelt. Das junge Weib näht, näht auf ihrem Lager, näht an einem  
<sup>10</sup> Stiefel. Er setzte sich neben das Mäusefräulein. Das Mäusefräulein  
 stand auf. Das Mäusefräulein bereitet für den Mann Essen, kocht

5

#### 4.4.2 Teckenkodning

UTF-8 är den huvudsakliga teckenkodningen i dagens (IT-)värld (Wikipedia 2014c):

*UTF-8 har valts som huvudsaklig teckenkodning i internetprotokoll: nya protokoll måste ge stöd till denna teckenkodning, om det inte av speciella skäl är olämpligt.*

De gränser som latin1 och andra äldre teckenkodningar erbjuder är godtyckliga. Med unicode har man samlat ihop "alla" tecken. Det finns naturligtvis mycket att göra (till exempel finns det problem med diakritiska tecken, för deras plats är inte alltid där den borde vara), men även ur lingvistisk synvinkel är läget bra, mycket bättre än någonsin förr.

Den normaliserade latin1 data i ISOs databaser räcker i princip mycket bra för forskare som kan sina kompetensområden och arbetar inom dessa ramar. Tyvärr är det inte alltid så. Kärnkunskaper bland forskare har mycket med svenskan och dess släktspråk att göra. Också finskan och de samiska språken utgör en bas för både de svenska och de finska och de samiska ortnamnen i Sverige. Man kan också tillfoga estniskan angående estlandssvenska ortnamn som källmaterial. När man har alla dessa källor (sannolikt kommer det någon gång finnas ytterligare källor), har man också tecken som inte ingår i latin1.

Nuförtiden har man lagliga skäl att forska i de samiska språken i Sverige. Och då är det svårt att undvika bokstäver utanför latin1. Antingen använder man ett fonetiskt alfabet med bokstäver säkert utanför latin1 eller så normaliserar man samiskan och då har man till exempel *Sverige* i nordsamiskan som *Ruotta*.

Man har med teckenkodning ett speciellt problem inom lingvistik i Sverige. Det berör sig just dialektologi, nämligen landsmålsalfabetet. Ortnamnen i Ortnamnsregistret är normaliserade, och innehåller inte landsmålsalfabetet, men ortnamnen på de ursprungliga sedeslapparna är ofta skrivna med landsmålsalfabetet. Landsmålsalfabetet är ett svenskt alfabet vars användning var och fortfarande är att skildra det svenska uttalet. Dessa tecken finns definitivt inte inom latin1 (figur 15). (Institutet för språk och folkminnen 2009.)

<b>1. Vokaler</b>		<i>ɲ:</i>	<i>rn</i> i rspr. <i>barn</i>
<i>a:</i>	i rspr. <i>tak</i>	<i>g:</i>	<i>ng</i> i rspr. <i>äng</i>
<i>ɑ:</i>	i rspr. <i>tack</i>	<i>r:</i>	rullande tungspets- <i>r</i>
<i>ä:</i>	mellanljud mellan <i>a</i> och <i>a</i>	<i>ʌ:</i>	skorrande <i>r</i>
<i>e:</i>	i rspr. <i>se</i>	<i>ʃ:</i>	<i>rs</i> i rspr. <i>fors</i>
<i>æ:</i>	mellanljud mellan <i>e</i> och <i>ä</i>	<i>ʂ:</i>	<i>tj</i> och <i>kj</i> i rspr. <i>tjäna, kjol</i>
<i>ɪ:</i>	i rspr. <i>lita, fick</i>	<i>f:</i>	<i>sj</i> i rspr. <i>sjö</i>
<i>o:</i>	i rspr. <i>bo, bott</i>	<i>ʃ:</i>	bakre <i>sje</i> -ljud
<i>u:</i>	i rspr. <i>gul</i>	<i>ʈ:</i>	<i>rt</i> i rspr. <i>hjort</i>
<i>ʊ:</i>	i rspr. <i>sund</i>		
<i>u:</i>	»bergslags- <i>u</i> »	<b>3. Särskilda tecken</b>	
<i>y:</i>	i rspr. <i>ny, nytt</i>	-	under ett tecken anger att ljudet är långt: <i>dag</i> ( <i>dag</i> ), <i>dag</i> ( <i>dagg</i> )
<i>ʏ:</i>	mellanljud mellan <i>y</i> och <i>ö</i>	-	under ett tecken anger att ljudet är halvlångt
<i>ɑ:</i>	i rspr. <i>gå</i>	-	under ett tecken anger att ljudet är stavelsebildande
<i>o:</i>	i rspr. <i>boll, håll</i>	-	under ett tecken anger att ljudet är svagt hörbart
<i>æ:</i>	i rspr. <i>räv</i>	-	anger att en tonande konsonant förlorat sin stämton ( <i>b</i> , <i>g</i> )
<i>a:</i>	i rspr. <i>här</i>	-	akut accent, t.ex. i rspr. <i>anden</i> (best. form av <i>and</i> )
<i>ø:</i>	i rspr. <i>söt</i>	-	grav accent, t.ex. i rspr. <i>anden</i> (best. form av <i>ande</i> )
<i>ø:</i>	i rspr. <i>sött</i>	-	starkt bitryck, t.ex. i <i>stènhus</i> ( <i>stenhus</i> )
<i>ø, ʉ:</i>	olika mellanljud mellan <i>ä</i> och <i>ö</i>		
<i>æ:</i>	obetonat <i>e</i> som i <i>pojke</i>		
<b>2. Konsonanter</b>			
<i>d:</i>	<i>rd</i> i rspr. <i>hård</i>		
<i>j:</i>	<i>j</i> -ljud		
<i>l:</i>	<i>rl</i> i rspr. <i>kärlek</i>		
<i>ʎ:</i>	»tjockt» <i>l</i> (även använt för <i>ʎ</i> )		

Figur 15. Landsmålsalfabetet (Institutet för språk och folkminnen 2009)

Trots allt finns det med UTF-8 som teckenkodning även idag problem i IT-världen. Kanske den allvarligaste saken för ISO/IEC JTC1 SC22 är att den standardteckenkodningen inom MySQL även idag är latin1, eller egentligen cp1252 (Oracle Corporation and/or its affiliates 2014b). Men man kan använda UTF-8: efter inloggningen i databasen ger man kommandot "SET NAMES 'utf8';" (Oracle Corporation and/or its affiliates 2014a).

Tekniskt sett är teckenkodningsproblemet redan en tickande tidsbomb. Det är så lätt att stava fel och att en bokstav utanför latin1 sannolikt kommer att dyka upp. Det finns en kolumn *atxt* i Ortnamnsregistret. Kolumnen är avsedd för ytterligare information. Det är svårt att använda bara latin1 här.

#### 4.4.3 Att korrigera fel

Det är alltid en god idé att kolla upp vad man har i en skriftlig form. Man kan dra nytta av generella program som gör en omfattande förstasteg ytlig textuellanalys, till exempel kan man ta reda på alla bokstäver som texten innehåller. När man har analyserat innehållet och hittat det som borde ändras, kan man modifiera och använda scripten att göra de ändringar som behövs. Sådana ändringar korrigerar mest stavfel.

Det finns program som gör textuella analyser. Och man kan programmera program eller hellre scripta små relativt enkla applikationer som kan modifieras för att hitta (potentiella) misstag. Som POC-applikation har man nu på IT-avdelningen en webbapplikation som analyserar texten och som slutsats ger de bokstäver man har i texten och deras enskilda antal, kanske en enkel början men sådana webbapplikationer har visat sig vara till hjälp.

#### 4.4.4 Loggande, när och var någonting har gjorts

Man kan automatiskt infoga användaren som har matat in data. Det går till på så sätt att man loggar in sig själv (i MySQL) och så är inmataren automatiskt tillfogad datainmatningen. Man kan också sköta loggandet med en (in och out)logfil. Loggandet är mer en administrations- och säkerhetsapplikation än något som har med forskningen att göra, och administrationen är troligen inte intresserad av saken. Därför är loggandet inte ett prioritetsprojekt, och blir sannolikt inte detta.

Man kan dock använda loggandet indirekt för andra orsaker. Man kan se mönstren med data, här kan datan hjälpa att även korrigera fel. Människor gör misstag det är oundvikligt. Man kan till exempel börja korrigera misstag i saker som gjorts på eftermiddagen (d.v.s. sannolikt trött och sannolikt mottagligare för att göra fel) och loggandet visar vilka dessa saker är. Här kan man till exempel använda webbapplikationen som söker bokstäver i texten och på så sätt hitta stavfel. I MySQL har man nog ett internt loggande med ändringstiderna. Användare kan i princip identifieras med IP:n. Det är alltså relativt enkelt att logga i

databasanvändningen, när man korrelerar användningen av MySQL och IP-aktiviteten med MySQL.



## 5 ÅTGÄRDSPLAN

### 5.1 Databasförändringar

Förändringarna är i allmänhet ganska stora och skulle kosta mycket.

Databasförändringsplanen är följande:

- Kolumn i databasen för det icke-normaliserade, dialektala ortnamnet
  - Ett långtidsprojekt, sannolikt kommer inte att realiseras
  - I stället gör detsamma, men i begränsade områden, såsom där älvdalskan talas
- En kolumn med en ytlig morfologisk analys gjord med datorlingvistiska program: kan förverkligas om forskarna någon gång har ett sådant behov
- En kolumn med etymologisk information
  - Bättre att ha en etymologisk databas, och kanske en kolumn i ortnamnsdatabasen som säger åtminstone att det finns en etymologi: en utvecklad POC-version finns på nätet; det är bättre att ha det här som ett sidoprojekt om behovet inte är stort just nu
- En runortnamnsdatabas är ett litet projekt som går att realiseras väldigt lätt

För kostnadens skull kommer stora förändringar kanske inte att realiseras. Man kan ändå göra små begränsade projekt.

#### 5.1.1 Att generera listor av ortnamn på olika skäl

Olika listor har alltid främjat forskning. Här är några listor som kan genereras och som är också värda att genereras. Åtminstone följande listor kan genereras:

- Baklängesordlistor
- Delade i förled och efterled
- Icke-delbara
- På semantisk grund

## 5.2 Mjukvaru- och uppdateringspolitik

Politiken i IT-världen är en svår sak. Därför måste man anpassa åtgärderna efter situationen. Denna lista kan hjälpa till med mjukvarupolitik:

- Kolla sakerna med kritiska programmens egna test suite, med testfall
- En enhetlig handling av diverse datakällor (Marshal): välj hellre öppna specifikationer än icke-öppna (XML/MySQL i stället för FileMaker, även .tex eller .dbk i stället för .docx)
- Uppdatera kritisk deprecated och end-of-life mjukvara (MySQL och PHP)
- Byta åldrande specifikationer (teckenkodning latin1 -> UTF-8 kanske med fontändringar och databasmotorn MyISAM -> InnoDB): men kolla läget noggrant
- Testfall som indikator
- POC-webbapplikation (som visar bokstäver i texten) och likadana (först speciellt för att korrigera stavfel)
- Ta reda på sökningens snabbhetsproblem
- Kritiska mjukvaran direkt av producenten: fungerar som den borde; bättre på att diagnosticera och korrigera fel

## 5.3 Inmatningsprogram

Det är verkligen lönsamt att använda XML med inmatningen. Formaten är i princip lätt och kan omformas. Sådana är alternativen:

- Skapa en XML-databas och generera XML-enhetsfiler, som forskarna matar in
- (Ett alternativ är att skapa en webbapplikation som liknar sökprogrammet, men detta alternativ är verkligen ett sämre alternativ)

## 5.4 Sökprogram

Här har man två utvecklingsmöjligheter som båda går att förverkliga där den ena betyder att man faktiskt inte gör något, eller låter saker göras automatiskt.

Möjligheterna är:

- Inga radikala förändringar
  - En ny uppdaterad POC-version är färdig att användas
  - En kolumn till vill forskarna definitivt ha i Ortnamnsregistret, nämligen beläggsår, och senare beläggsår: gjort med OCR

och

- Utveckla ett nytt sökprogram med en karta-GUI i fokus: använda vad man kan från det nuvarande sökprogrammet

## 6 SAMMANFATTNING

Detta arbete har försökt kasta ljus över de områden som kan ändras eller i allmänhet utvecklas i Ortnamnsregistret med hjälp av IT-avdelningen.

Forskarna tycks vara ganska nöjda med det som IT-avdelningen kan erbjuda. De forskningsbehov som har med databaser att göra och som ortnamnsforskarna kan föreslå är komplicerade projekt som ger små resultat men kräver väldigt mycket arbete. Några saker, till exempel uppdateringar och fonter, kan vara problematiska men problemen är troligen inte så allvarliga nu. Ett litet större problem är ett inmatningsprogram för (sydsvenska) ortnamn. Om man bestämmer att använda en XML-databas, så är det kanske bättre att generera lämpliga XML-enhetsfiler så att forskare kan fylla in data där med diverse XML-WYSIWYG-program (eller bara med normala editorer).

En sak som kan främja forskningen på ISOF är att IT-avdelningen kan generera något som en början på forskning. Det finns kriterier för att hitta åtminstone möjliga fall som kan innehålla något viktigt, till exempel ovanliga ortnamnsefterleder. Problemet är att sådana fall är svåra eller omöjliga att skildra formellt. Här kan man alltså börja ta ut de fall som troligen inte är viktiga, och sedan gå igenom det som blir kvar. Andra fall kan ha något med fonotax att göra. Då kan man ge till exempel en lista över bokstavsträngar att hitta i ortnamnen. Semantiska särdrag är svårare men något kanske kan uppnås med tanke på morfologin. Saken är alltså så att forskare har några (oformella) idéer om potentiellt intressanta ortnamnsfall som man kan hitta i databaser, eller åtminstone kasta bort de fall som inte alls är intressanta. Men man måste skildra det som behövs noggrant.

En överraskning var att underhållet av själva databaserna och databasapplikationerna tycks behöva förändringar. Med uppdatering är några saker ”bara” gammalmodiga, men några måste ändras med nya versioner av mjukvaran om applikationerna ska fungera överhuvudtaget. Förändringarna blir kanske lite svåra eftersom serverna tas hand om av en firma och inte av IT-avdelningen. Ända fungerar allt och man måste själv utvärdera om det lönar sig att uppdatera och ta till användning nya versioner av mjukvaran. Men när

omorganisationen med databaser börjar under hösten, är det nog ändå bra att förbereda sig för framtiden och göra förändringar och därför har vi här en åtgärdsplan.

## BIBLIOGRAFI

Ainiala, T. & Pitkänen, R. L. 2002. Onomastic research in Finland. Nyirkos, István (ed.). History of the study of toponyms in the Uralian languages. *Onomastica Uralica* 2, 81–119. Debrecen–Hki.  
([http://onomaurl.klte.hu/onomural/gb/frame\\_gb1.html](http://onomaurl.klte.hu/onomural/gb/frame_gb1.html))

Bradford, R. 2011. Five reasons to upgrade to MySQL 5.5. Bradford, R. [citerat 31.3.2014]. länken: <http://ronaldbradford.com/blog/five-reasons-to-upgrade-to-mysql-5-5-2010-12-15>

Institutet för språk och folkminnen. 2009. Uttalsbeteckningar enligt landsmålsalfabetet. Institutet för språk och folkminnen. [citerat 8.4.2014]. länken: [http://www.sofi.se/GetDoc?meta\\_id=1344](http://www.sofi.se/GetDoc?meta_id=1344)

Institutet för språk och folkminnen. 2014. Institutet för språk och folkminnen. Institutet för språk och folkminnen. [citerat 26.3.2014]. länken: <http://www.sofi.se/>

ioMeWeekly. 2013. Benchmark PHP mysql vs PHP mysqli. [iomeweekly.blogspot.sg](http://iomeweekly.blogspot.sg). [citerat 31.3.2014]. länken: <http://iomeweekly.blogspot.sg/2013/04/benchmark-php-mysql-vs-php-mysqli.html>

Launchpad. 2012. Bug report on mysql5.5, Ubuntu 12.04. Canonical Ltd. [citerat 2.4.2014]. länken: <https://bugs.launchpad.net/ubuntu/+source/mysql-5.5/+bug/959683>

Lehtisalo, T. 1947. Juraksamojedische Volksdichtung. Suomalais-ugrilaisen seuran toimituksia XC. Helsinki: Suomalais-ugrilainen seura.

Lourdass, V. 2013. PHP 5.3 vs 5.4 vs 5.5: Speed and memory usage. [lourdass.name](http://www.lourdass.name). [citerat 30.3.2014]. länken: <http://www.lourdass.name/blog/php-53-vs-54-vs-55-speed-and-memory-usage>

Morottaja, M. 2006. Language. Sámi musea Siida & Anarâškielâ servi. [citerat 30.3.2014]. länken: <http://www.samimuseum.fi/anaras/english/kieli/kieli.html>

Namnarkivet i Uppsala. 2010. Ortnamnsarkivet i Uppsala. Namnarkivet i Uppsala. [citerat 30.3.2014]. länken: <http://www.sofi.se/1319>

Namnarkivet i Uppsala. 2014. Ortnamnsregistret. Namnarkivet i Uppsala. [citerat 26.3.2014]. länken: <http://www.sofi.se/ortnamnsregistret>

Oracle Corporation and/or its affiliates. 2014a. MySQL 5.5 Reference Manual: Chapter 10.1 Character Set Support. Oracle Corporation and/or its affiliates. [citerat 6.4.2014]. länken: <http://dev.mysql.com/doc/refman/5.5/en/charset.html>

Oracle Corporation and/or its affiliates. 2014b. MySQL 5.5 Reference Manual: Chapter 10.1.14.2 West European Character Sets. Oracle Corporation and/or its affiliates. [citerat 6.4.2014]. länken: <http://dev.mysql.com/doc/refman/5.5/en/charset-we-sets.html>

Oracle Corporation and/or its affiliates. 2014c. The MySQL Test Framework, Version 2.0: Chapter 5 MySQL Test Programs. [citerat 1.4.2014]. länken: <http://dev.mysql.com/doc/mysqltest/2.0/en/test-programs.html>

Peterson, L. 2002. Nordiskt runnamnslexikon. Fjärde, reviderade versionen med tillägg av frekvenstabeller och finalalfabetisk namnlista Februari 2002. Institutet för språk och folkminnen. [citerat 3.4.2014]. länken: [http://www.sofi.se/servlet/GetDoc?meta\\_id=1472](http://www.sofi.se/servlet/GetDoc?meta_id=1472)

R2B Software. 2014. Marshal. R2B Software. [citerat 2.4.2014]. länken: [http://www.risetobloome.com/Page\\_1\\_S\\_NodeListing.aspx?item=1830](http://www.risetobloome.com/Page_1_S_NodeListing.aspx?item=1830)

Rédei, K. 1988–1991. Uralisches etymologisches Wörterbuch. Wiesbaden: Otto Harrassowitz.

Schwartz, B. 2011. What's the recommended MySQL version? Percona LLC. [citerat 31.3.2014]. länken: <http://www.mysqlperformanceblog.com/2011/08/15/whats-the-recommended-mysql-version>

Schwartz, B., Zaitsev, P. och Tkachenko, V. 2012 High Performance MySQL. 3. upplaga. Cambridge: O'Reilly.

Sigvardsson, S. 2014. Dokumentation av Ortnamnsregistret: Beskrivning av tabellstruktur och datainsättning. Institutet för språk och folkminnen. Ett opublicerat dokument.

SitePoint. 2013. Forum: STOP using mysql - migrating to procedural mysqli.

SitePoint. [citerat 31.3.2014]. länken:

<http://www.sitepoint.com/forums/showthread.php?1182768-STOP-using-mysql-migrating-to-procedural-mysqli>

Svanberg, I. 1986. Sockenlappar. Tidskrift utgiven av föreningen för svensk kulturhistoria i samarbete med nordiska museet och folklivsarkivet i Lund. 69:4. 97–117.

Svenska Akademien. 2006. Svenska Akademiens ordlista över svenska språket.

Svenska Akademien. [citerat 6.4.2014]. länken:

[http://www.svenskaakademien.se/svenska\\_spraket/svenska\\_akademiens\\_ordlista/saol\\_pa\\_natet](http://www.svenskaakademien.se/svenska_spraket/svenska_akademiens_ordlista/saol_pa_natet)

Treworky, J. 2014. ImageMapster. <http://www.outsharked.com/>. [citerat

12.4.2014]. länken: <http://www.outsharked.com/imagemapster/>

The PHP Group. 2014a. import\_request\_variables. The PHP Group. [citerat

30.3.2014]. länken: <http://www.php.net/manual/en/function.import-request-variables.php>

The PHP Group. 2014b. Unsupported Branches. The PHP Group. [citerat

27.3.2014]. länken: <http://php.net/eol.php>

Universität Koblenz-Landau. 2014. Uralothek. Universität Koblenz-Landau.

[citerat 26.3.2014]. länken: <http://uralothek.uni-koblenz.de:8080/Uralothek/pdom/basis.html>

Vikstrand, P. 2008. Folketymologier. Om ortnamnens meningsskapande funktion.

Guðrún Kvaran, Hallgrímur J. Ámundason, Jónína Hafsteinsdóttir & Svavar

Sigmundsson (red.) Norræn nöfn – Nöfn á Norðurlöndum. Hefðir og endurnýun.

Nordiska namn – Namn i Norden. Tradition och förnyelse. Handlingar från Den



fjortonde nordiska namnforskarkongressen i Borgarnes 11–14 augusti 2007. (NORNA-rapporter 84.) Uppsala 2008. S. 461–468. (<http://www.sofi.se/5559>)

Vikstrand, P. 2013. NAU:s kravspecifikation för de datoriserade ortnamnssamlingarna. Namnarkivet hos Institutet för språk och folkminnen, Uppsala. Ett opublicerat dokument.

Wikipedia. 2014a. MySQL. Wikipedia. [citerat 7.5.2014]. länken: <http://en.wikipedia.org/wiki/Mysql>

Wikipedia. 2014b. PHP. Wikipedia. [citerat 7.5.2014]. länken: <http://en.wikipedia.org/wiki/PHP>

Wikipedia. 2014c. UTF-8. Wikipedia. [citerat 26.3.2014]. länken: <http://sv.wikipedia.org/wiki/UTF-8>

Wise, M. 2010. ORD-kurs 2. Ortnamn, runor, dialekter. Runor. (Kursbeskrivning). Monica Wise. [citerat 30.3.2014]. länken: <https://www.liu.se/ikk/svspr/ord-kurs-2/?l=sv>